



Defence Research and
Development Canada

Recherche et développement
pour la défense Canada



Using Profiles in GOSSIP to Examine Concepts Associated with the Afghan National Police in Open Source Media

*Peter Kwantes
DRDC Toronto*

*Ian Lawless
TFK Psychological Operations*

Defence R&D Canada
Technical Memorandum
DRDC Toronto TM 2011-085
October 2011

Canada

Using Profiles in GOSSIP to Examine Concepts Associated with the Afghan National Police in Open Source Media

Peter Kwantes
DRDC Toronto

Ian Lawless
TFK Psychological Operations

Defence R&D Canada – Toronto

Technical Memorandum
DRDC Toronto TM 2011-085
October 2011

Principal Author

Original signed by Peter Kwantes

Peter Kwantes

Defence Scientist

Approved by

Original signed by Keith Stewart

Keith Stewart

Head, Socio-Cognitive Systems Section

Approved for release by

Original signed by Dr. Stergios Stergiopoulos

Dr. Stergios Stergiopoulos

Acting Chair, Knowledge and Information Management Committee

Acting Chief Scientist

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2011

© Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2011

Abstract

The Graphical Overview of the Social and Semantic Interactions of People (GOSSIP) is a software tool developed by Defence Research and Development Canada – Toronto (DRDC Toronto). The program is designed to help the operator/analyst develop a fast and accurate impression of the relationships among entities (people, places, organizations) discussed in document collections that are too large to read through in a reasonable amount of time. Over the past few years, North Atlantic Treaty Organization (NATO) countries have developed and delivered a training program for Afghan National Police (ANP) members. The ANP are considered by some to be unprofessional, inept, corrupt, and as having little positive effect on the local population's sense of security in their communities. These qualities of the ANP were explored using GOSSIP by examining an open source media collection prepared for analysts and commanders in Kandahar Airfield (KAF). We found that when the ANP was discussed in the media updates provided to Task Force Kandahar (TFK) commanders, it was very often in a positive way. In particular, discussion about NATO's role in professionalizing the ANP dominated articles about the ANP. We propose that the extent to which the ANP is discussed in positively toned articles might lead the target audience for these articles to have an unduly positive impression of the ANP. GOSSIP is a prototype. It needs to be developed further to allow it to be used as a web-based device on a network. Future work should also enhance the tool by providing it with the ability to scrape information from various sources without the user having to load documents manually.

Résumé

Recherche et développement pour la défense Canada – Toronto (RDDC Toronto) a conçu l'outil logiciel d'aperçu graphique de l'interaction sociale et sémantique entre les personnes (GOSSIP). Ce programme permet à l'opérateur ou l'analyste d'avoir un aperçu rapide et exact des dynamiques interrelationnelles (personnes, places, organismes) présentées dans des documents ne pouvant être lus dans des délais raisonnables en raison de leur volume. Au cours des dernières années, les pays membres de l'Organisation du Traité de l'Atlantique Nord (OTAN) ont mis sur pied un programme d'instruction pour la Police nationale afghane (PNA). Certaines personnes estiment que les membres de la PNA manquent de professionnalisme en plus d'être déplacés et corrompus, et qu'ils apportent un faible sentiment de sécurité à la population locale. Les analystes et commandants de l'aérodrome de Kandahar (KAF) ont examiné, à l'aide du GOSSIP, ces particularités à partir d'une médiathèque ouverte constituée spécialement pour eux. Or, il s'avère que les dernières nouvelles médiatiques sur la PNA qui ont été remises aux commandants de la Force opérationnelle de Kandahar (FO Kandahar) étaient très positives. Les articles traitaient notamment du rôle de l'OTAN en ce qui a trait au professionnalisme de la police afghane. Ainsi, des articles favorables à la PNA pourraient amener le public cible à avoir une meilleure opinion d'elle. Le GOSSIP est un prototype. Il faut le développer davantage afin qu'il devienne un outil Web pouvant être utilisé sur un réseau. Avec le temps, il pourra également regrouper des données de diverses sources sans que l'utilisateur ait à télécharger les documents manuellement.

This page intentionally left blank.

Executive summary

Using Profiles in GOSSIP to Examine Concepts Associated with the ANP in Open Source Media

Peter Kwantes, Ian Lawless; DRDC Toronto TM 2011-085M 2011-085; Defence R&D Canada – TorontoToronto;

Introduction or background: The Graphical Overview of the Social and Semantic Interactions of People (GOSSIP) is a software tool developed by Defence Research and Development Canada – Toronto (DRDC Toronto). The program is designed to help the operator/analyst develop a fast and accurate impression of the relationships among entities (people, places, organizations) discussed in document collections that are too large to read through in a reasonable amount of time. Over the past few years, North Atlantic Treaty Organization (NATO) countries have developed and delivered a training program for Afghan National Police (ANP) members. The ANP are considered by some to be unprofessional, inept, corrupt, and as having little positive effect on the local population's sense of security in their communities. These qualities of the ANP were explored using GOSSIP by examining an open source media collection prepared for analysts and commanders in Kandahar Airfield.

Results: We found that when the ANP was discussed in the media updates provided to Task Force Kandahar commanders, it was very often in a positive way. In particular, discussion about NATO's role in professionalizing the ANP dominated articles about the ANP.

Significance: We propose that the extent to which the ANP is discussed in positively toned articles might lead the target audience for these articles to have an unduly positive impression of the ANP.

Future plans: GOSSIP is a prototype. It needs to be developed further to allow it to be used as a web-based device on a network. Future work should also enhance the tool by providing it with the ability to scrape information from various sources without the user having to load documents manually.

Sommaire

Utilisation de profils dans le GOSSIP afin d'examiner les concepts reliés à la Police nationale afghane dans les médias :

Peter Kwantes, Ian Lawless; DRDC Toronto TM 2011-085; R & D pour la défense Canada – Toronto.

Introduction ou contexte : Recherche et développement pour la défense Canada – Toronto (RDDC Toronto) a conçu l'outil logiciel d'aperçu graphique de l'interaction sociale et sémantique entre les personnes (GOSSIP). Ce programme permet à l'opérateur ou l'analyste d'avoir un aperçu rapide et exact des dynamiques interrelationnelles (personnes, places, organismes) présentées dans des documents ne pouvant être lus dans des délais raisonnables en raison de leur volume. Au cours des dernières années, les pays membres de l'Organisation du Traité de l'Atlantique Nord (OTAN) ont mis sur pied un programme d'instruction pour la Police nationale afghane (PNA). Certaines personnes estiment que les membres de la PNA manquent de professionnalisme en plus d'être déplacés et corrompus, et qu'ils apportent un faible sentiment de sécurité à la population locale. Les analystes et commandants de l'aérodrome de Kandahar (KAF) ont examiné, à l'aide du GOSSIP, ces particularités à partir d'une médiathèque ouverte constituée spécialement pour eux.

Résultats : Or, il s'avère que les dernières nouvelles médiatiques sur la PNA qui ont été remises aux commandants de la Force opérationnelle de Kandahar (FO Kandahar) étaient très positives. Les articles traitaient notamment du rôle de l'OTAN en ce qui a trait au professionnalisme de la police afghane.

Portée : Ainsi, des articles favorables à la PNA pourraient amener le public cible à avoir une meilleure opinion d'elle.

Perspectives: Le GOSSIP est un prototype. Il faut le développer davantage afin qu'il devienne un outil Web pouvant être utilisé sur un réseau. Avec le temps, il pourra également regrouper des données de diverses sources sans que l'utilisateur ait à télécharger les documents manuellement.

Table of contents

| | |
|---|-----|
| Abstract | i |
| Résumé | i |
| Executive summary | iii |
| Sommaire | iv |
| Table of contents | v |
| List of figures | vi |
| List of tables | vi |
| Acknowledgements | vii |
| Introduction | 1 |
| Profiling in GOSSIP | 2 |
| The Current Study..... | 3 |
| Method..... | 3 |
| Results..... | 4 |
| Discussion..... | 5 |
| What do the results mean?..... | 7 |
| What potential impact could these associations with concepts have on the reader's impression of entities like the ANP? | 7 |
| How well did GOSSIP do? | 8 |
| Future Directions for GOSSIP as a Tool for Influence Operations | 9 |
| Conclusions and Recommendations | 10 |
| References | 11 |
| List of symbols/abbreviations/acronyms/initialisms | 13 |
| Distribution list..... | 14 |

List of figures

Figure 1. GOSSIP's user interface. In this screenshot, the user can view the connections possessed by Brad Pitt and Angelina Jolie in documents scraped from the internet movie database (IMDB) ... 1

Figure 2. In the bottom row, the concept for CORRUPTION is created by summing the columns formed by placing the vectors of the defining terms atop one another2

Figure 3. Bar chart depicting the association between the ANP and each of the concepts under consideration.....5

List of tables

Table I: Concepts used in the study and their defining terms. Note: the concept of PROFESSIONALISM included both the American and British spellings of the affixed forms of “honour”. *OMLT = Operational Mentor and Liaison Team 4

Acknowledgements

The work conducted for this report was done under ARP 15ah02 as part of a Technical Assistance Visit (TAV) to Kandahar Airfield (KAF) between May 16 and May 26, 2011 in support of CF PSYOPS Element of Task Force Afghanistan, Roto 3-10. We gratefully acknowledge the contributions of Maj. Benoit Mainville, CO of Roto 3-10's Psychological Operations platoon in KAF for requesting the TAV and providing analysts to help conduct the work reported here. We also acknowledge the assistance of MCpl. Ryan Blaschuk at RANA Radio in Kingston, Ontario who helped by providing us with the media updates required for the content analysis we conducted.

This page intentionally left blank.

Introduction

The Graphical Overview of the Social and Semantic Interactions of People (GOSSIP, Kwantes & Terhaar, 2010) is a software tool developed by Defence Research and Development Canada – Toronto (DRDC Toronto). The program is designed to help the operator/analyst develop a fast and accurate impression of the relationships among entities (people, places, organizations) discussed in document collections that are too large to read through in a reasonable amount of time. The program was designed for the situation in which an analyst might have to go through tens of thousands of documents from a domain to learn about who are the influential people and organizations.

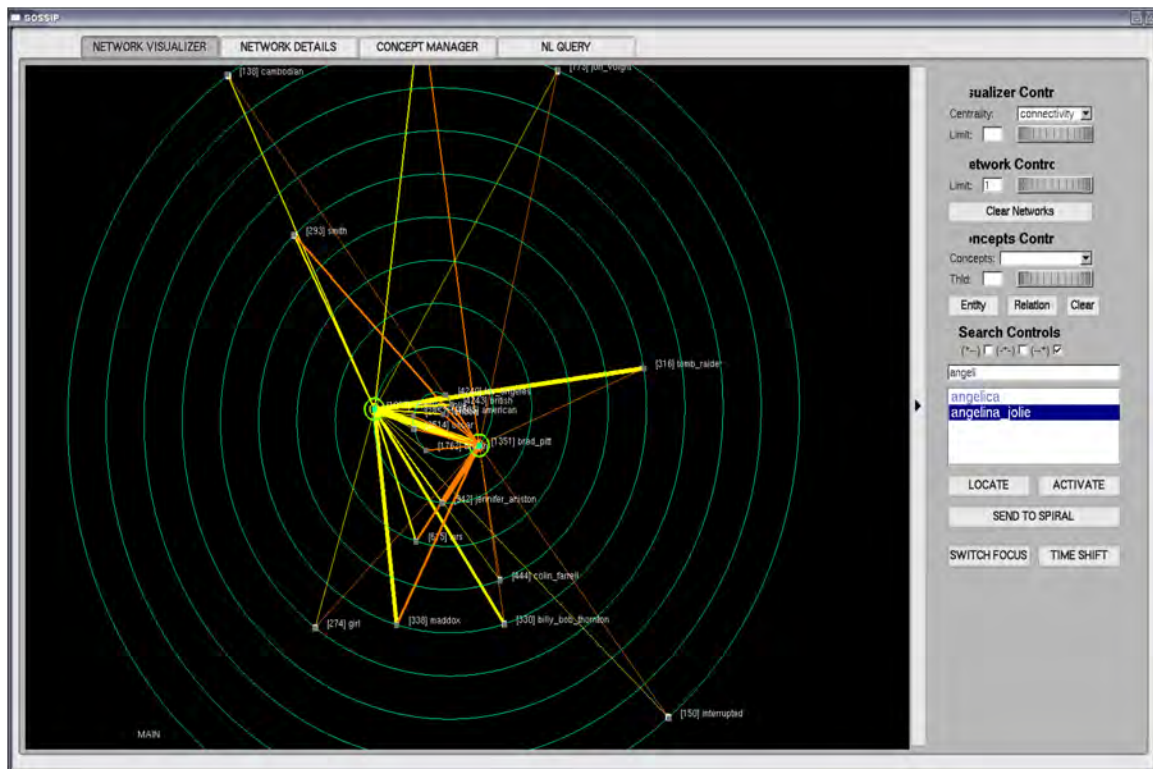


Figure 1: GOSSIP's user interface. In this screenshot, the user can view the connections possessed by Brad Pitt and Angelina Jolie in documents scraped from the internet movie database (IMDB)

GOSSIP is a visualization tool that allows the user to see the connections that exist among entities (see *Figure 1*). A “connection” in GOSSIP refers to the co-occurrence of entities in the same document. In addition to seeing co-occurrence information, GOSSIP gives the user an indication of the importance the entities play in the domain covered by the documents. The “importance” of

an entity in this context refers to the number of connections enjoyed by an entity. From the visualization interface, the user can drill down to the relevant source documents to determine the precise nature of the connections found by the system. Apart from the clerical information regarding the presence and strength of connections, GOSSIP has a computational model of semantics running in the background that processes all of the documents to create “meaning” representations for every word it encounters in the collection. Without going into detail, the model, named Latent Semantic Analysis (LSA), reads the documents, and in a completely unsupervised fashion generates a semantic representation for every content word and entity found in the collection. The representation takes the form of a large vector. A vector for one term can be compared to that of another by measuring the cosine between the two. A cosine provides a value much like that of a correlation coefficient in that, a cosine of 1 indicates that two vectors are identical, and a cosine of 0 indicates that they are orthogonal.

Profiling in GOSSIP

Having GOSSIP develop its own semantic representations for the materials in a document collection provides the user with the ability to extract meaning-based information from the documents. In this report, we will focus on GOSSIP's capability to generate a profile of an entity based on a set of user defined concepts. For example, a user may be interested to know how associated someone like Prime Minister Stephen Harper is to the concepts of LEADERSHIP, FAMILY, MUSIC and CORRUPTION in the collection of documents that discusses him. It is important to note, that an entity's association to a concept does not have any value judgment associated with it. In other words, if Stephen Harper has a strong association to the concept of CORRUPTION in my documents, it does not mean that he is corrupt. It simply means he has a strong association to the concept, which might exist because our document collection discusses him as a fighter of corruption.

| | | | | | | |
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| bribery | .2 | .1 | .1 | .7 | .2 | .1 |
| corrupt | .3 | .3 | .1 | .2 | .1 | .3 |
| blackmail | .1 | 0 | .4 | 0 | .2 | .1 |
| CORRUPTION | .6 | .4 | .6 | .9 | .5 | .5 |

Figure 2: In the bottom row, the concept for CORRUPTION is created by summing the columns formed by placing the vectors of the defining terms atop one another.

For GOSSIP, a concept is a collection of words that, together, define an idea. So, for example, the concept of COURRUPTION might be defined by a collection of terms including, *bribery*, *corrupt*, *blackmail*, and so on. To create the concept for CORRUPTION in GOSSIP, we sum the LSA semantic vectors for the defining words to create a single vector for the concept. *Figure 2* contains an example of vector addition for the topic, CORRUPTION. Each row of the figure contains a vector for a term that we consider a member of the concept, CORRUPTION. To sum

them, we align the rows atop one another, and sum the columns to create a new, concept vector. The concept vector's similarity to an entity's vector is expressed as the cosine between them.

The Current Study

The Afghan National Police (ANP) is a government-funded police organization manned by personnel many of whom are inexperienced, underage with high illiteracy rates. Over the past few years, North Atlantic Treaty Organization (NATO) countries have developed and delivered a training program for ANP members. The ANP are considered by some to be unprofessional, inept, corrupt, and as having little positive effect on the local population's sense of security in their communities. These qualities of the ANP were explored using GOSSIP by examining an open source media collection prepared for analysts and commanders in Kandahar Airfield (KAF).

Thus far, GOSSIP has never been used in an operational context. In this report, we describe how it was used to provide support to the Psychological Operations (PSYOPS) element of Task Force Kandahar (TFK) Rotation 3-10 in KAF. Specifically, the program was used to help uncover how the open source media discusses the ANP and how those discussions might shape readers' perceptions of that organization. The concepts of particular interest to PSYOPS analysts were the extent to which the ANP:

- were associated with the notion of safety and security among locals,
- were effective in their job,
- conducted themselves in a professional manner.
- were discussed with respect to corruption, and
- were discussed with respect to training provided by NATO.

Method

Materials. The open source media collection used for this study consisted of approximately two year's (2009 – 2011) worth of news articles from Afghan and International sources. The Afghan local news stories were written in the local languages and translated into English for the International Security Assistance Force (ISAF) personnel. The particular collection we worked with in this study came from daily media updates provided for commanders and analysts, and comprised approximately 11,000 articles.

Procedure. The collection of documents was loaded into GOSSIP for processing. A semantic space was derived for the terms in the collection using LSA. We created five concepts for GOSSIP against which we evaluated the ANP. The concepts were: SAFETY, EFFECTIVENESS, PROFESSIONALISM, CORRUPTION and TRAINING. The terms used to define each concept are listed in *Table 1*. (Note that throughout this document, we will spell concept names using upper case letters when discussing them as GOSSIP's vector representation of them.)

Table I: Concepts used in the study and their defining terms. Note: the concept of PROFESSIONALISM included both the American and British spellings of the affixed forms of “honour”. *OMLT = Operational Mentor and Liaison Team

| CONCEPT | TERMS |
|-----------------|---|
| SAFETY | safe safety secure |
| EFFECTIVENESS | effectiveness effective successful success |
| PROFESSIONALISM | professional professionalism trust honest decency honesty honorable honourable honor honour integrity |
| CORRUPTION | corrupt bribe corruption bribery intimidation intimidate beat beating harass harassment |
| TRAINING | trained training train mentor mentoring OMLT* |

As mentioned above, GOSSIP has the capability to provide the user with a profile of an entity across any number of constructed concepts. As a next step then, GOSSIP generated a profile of the ANP as it relates to the five concepts above. As an initial processing step, GOSSIP calculates the similarity between the vectors describing an entity and a concept by calculating the vector cosine between the two. In a second step, the program calculates the extent to which the relationship between the entity and concept is greater than what would be expected from that of a randomly selected entity. More specifically, GOSSIP randomly samples 2000 entities and terms from its database and calculates the cosine between each and the concept. From this, it calculates the mean and standard deviation of the cosines in the sample. The mean and standard deviation are then used to re-express our entity's relationship to the concept in terms of how much more associated it is to the concept than “the average” entity or term mentioned in the collection. For example, the similarity between the ANP and the concept of AFGHANISTAN would be very high. However, the strong relationship does not indicate an association that is unique to the ANP. Many entities and terms in the corpus would have a strong association with AFGHANISTAN. What we would want to know is whether the ANP's association with the concept is substantially higher than a baseline that we would expect in the general population of entities and terms. In other words, to what extent is the association between an entity and a concept unique to that concept? To provide some added context, a score of 1.0, 2.0 and 3.0 represent associations that are stronger than 84%, 95% and 99% of the associations between the concept and the entities in our sampled distribution. A score of 0.0 associated with an entity means that the entity's association with the concept sits at the 50th percentile. That is, half the entities and terms in our random sample have associations with this concept that are as high or higher as our entity of interest. Re-expressing the similarities as a normalized deviation (a Z-score) from the average entity's association with the concept captures the information we need from the analysis.

Results

The relationship between the vector for the entity, ANP and each of the five concepts we defined above are shown in *Figure 3*. As is clear in the figure, documents that discuss the ANP do not seem to be any more associated to the concepts of SAFETY and EFFECTIVENESS than any of

the other terms and entities sampled in the distribution. It is worth noting again that a low value close to zero for a concept does not necessarily mean that the entity has a low association with a concept; it could be a very strong association. **A low value simply means that the entity's association to a concept is no stronger than the association that many other entities and words in the documents have to the same concept.**

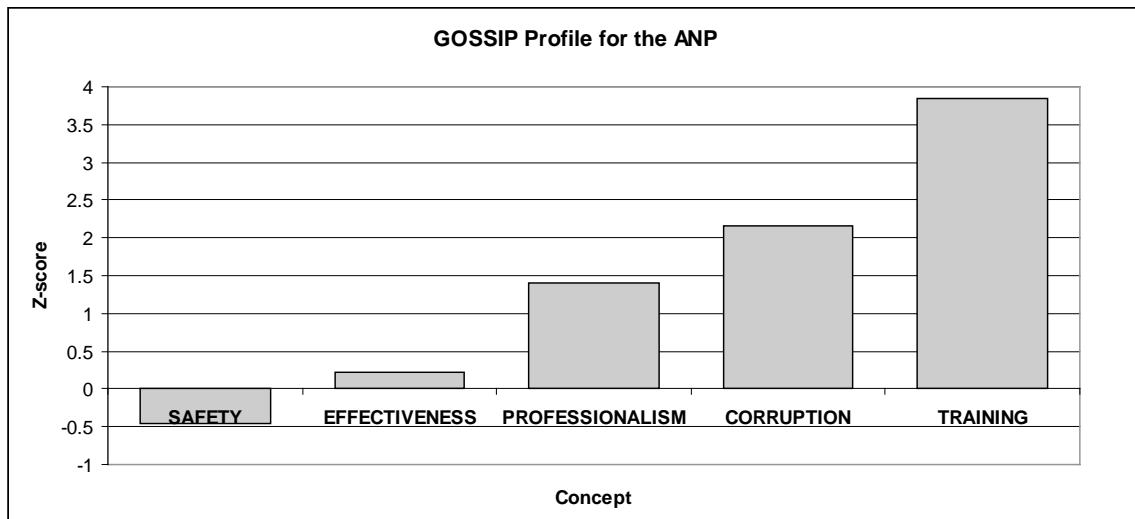


Figure 3: Bar chart depicting the association between the ANP and each of the concepts under consideration.

On the other hand, the associations between the ANP and the remaining concepts are noteworthy. In particular, the association between the ANP and the concepts of PROFESSIONALISM and CORRUPTION was higher than approximately 97% of the cosines in the sampled distribution. The most striking association was between the ANP and the concept of TRAINING. For this concept, the relationship was stronger than nearly 100% of the associations in the sampled terms and entities.

Discussion

In this study, we generated five concepts, and compared their vectors to the vector representing the ANP. We found that the qualities OF PROFESSIONALISM, CORRUPTION, and TRAINING were salient in the document collection provided to us. The ANP's association to the qualities OF SAFETY and EFFECTIVENESS were no stronger than approximately 50% of the entities and terms in our sample.

The presence of an association between an entity and a concept tells the user nothing about the nature of the association. For example, the ANP's strong association to the concept, CORRUPTION may exist because they are seen as corrupt or because the documents discuss them as being fighters of corruption. All the user can know, from the profile, is what concepts are salient in the collection under examination. In what follows, we will discuss our examination of the documents of our collection mentioning the ANP (186 in total) and assess how the concepts are discussed. It is worth noting that our assessment of the documents may be more cursory than might be ideal. The assessment entailed performing multiple term searches on the document to

find instances of words relevant to each of the concepts, tallying the number of documents that were relevant to our query, and noting their dominant themes. It should be mentioned, that under normal circumstances, users would likely ignore exploring the documents in which the association score was very low. For the purposes of this demonstration however, and for the sake of thoroughness, we will examine the 186 ANP-related documents for each concept in our profile.

Safety. The purpose of including a concept for SAFETY was to determine if the work or presence of the ANP provided any sense of security to Afghans. To do so, we searched for the terms *safe* and *safety* in the documents. There were 18 documents that mentioned the terms, of which three articles made mention of Afghans feeling safe and secure because of the ANP. The conclusion to be drawn from the assessment of the documents is straightforward: Afghans are concerned about their safety, and some have mentioned that their sense of security is increased from the ANP's presence. However, the issue of safety is pervasive in the document collection, and not limited to discussions about the ANP.

Effectiveness. To determine the extent to which the ANP is perceived to be effective by Afghans, we searched the 186 documents for instances of the terms *effective*, *effectiveness*, *success*, and *successful*. The collection contained a total of 28 documents mentioning or discussing effectiveness. Nineteen of those documents were not discussions about the ANP *per se*, but rather the effectiveness of the NATO-led training program set in place for them. Nine of the 28 documents discussed the effectiveness of the ANP, of which three were positive in tone. Again, effectiveness is a common theme in the documents. The ANP is not the only organization whose effectiveness is followed and reported by the media. The Afghan National Army (ANA), Afghan Border Police (ABP), ISAF are also organizations whose effectiveness is discussed in the documents.

Professionalism. This concept loaded heavily on the ANP in our profile. A search in the documents for the terms, *professional(ism)*, *integrity*, *honest(y)* and *trust* turned up 18 documents. Of those documents, seven discussed the ANP's lack of professionalism. The remaining documents highlighted NATO's training to build a professional police force and army. It is worth pointing out that despite there being as many or fewer documents discussing PROFESSIONALISM (18) than documents discussing either EFFECTIVENESS (28) or SAFETY (18), the salience of the relationship between this concept and the ANP was high in the sense that, as we have defined the concept, the association between the ANP and PROFESSIONALISM is substantially higher than the average entity's association with it.

Corruption. This concept also loaded heavily on the ANP with a score indicating that the ANP's association with CORRUPTION was higher than over 95% of the entities in our sample. Again, the strong score tells the user that, to a great degree, the ANP has a fairly exclusive association with the concept. Twenty-three of the 186 documents mentioning the ANP included terms like, *corrupt(ion)*, *bribe(ry)*, *beat(ing)*, and *harass(ment)*. Of them, 12 documents discuss or mention a problem of corruption in the ANP. The remaining documents discuss more general problems of corruption in government, locals beating locals, and sexual harassment within the military ranks.

Training. The concept of TRAINING was the dominant concept related to the ANP. More specifically, the ANP's association with TRAINING was higher than the associations with any of the entities sampled for our baseline. Put another way, when the concept of TRAINING is discussed in this collection, it was very often in the context of the ANP. Of the 54 documents

mentioning training in the 186 containing the ANP, 46 were on the topic of NATO's and the Canadian Forces' (CF) role in training the ANP and its effectiveness. The remaining 8 documents were on training topics that were not specifically related to the ANP's training.

What do the results mean?

The analysis above paints a clear picture of how exclusively associated the concepts we have created are associated with the ANP. Clearly, topics like the locals' sense of security (SAFETY) and effectiveness are salient among the documents. However, they are not topics uniquely associated with the ANP. Indeed, they are salient to probably most entities and terms mentioned in the collection. With respect to the ANP however, it appears clear that discussions of corruption, professionalism, and training are reserved heavily, and sometimes almost exclusively for them in the documents. It is worth noting, however, that for the topic of training, the ANA is often mentioned alongside the ANP. Therefore, while the ANP's relationship to TRAINING is strong, it is in many documents, an association shared with the ANA.

What potential impact could these associations with concepts have on the reader's impression of entities like the ANP?

In psychology, there are a number of well-studied cognitive biases that affect decision-making. Among them, the availability heuristic has potential relevance here. The *Availability Heuristic*, refers to the tendency for people to predict the frequency of an event based on how easily the event comes to mind (Tversky & Kahneman, 1973). For example, imagine being asked to come up with the capital city names for the following countries: Brazil, Australia, and Canada. Many people would answer, Rio de Janeiro, Sydney, and Toronto because they are the names that come most quickly to mind when they think of the country name. All three answers are incorrect; the capitals are Brasilia, Canberra, and Ottawa, respectively. Even if they have learned the correct city name in the past, people can still report the wrong city name because we are generally biased to report that which comes most easily to mind.

The documents in the collection under examination here, could lend themselves to a similar kind of bias with respect to the impressions generated for entities discussed in their documents. It is worth mentioning, that the ideas in the following discussion are speculative. Published work on the *Availability Heuristic* is typically conducted within the context of decision-making research. In this report, we are extending the basic ideas to the associations a reader forms as s/he reads a collection of documents on a particular topic.

We postulate that the relationships between an entity and its associated concepts can serve to shape the reader's impression of the entity. Entities have associations to many concepts in a given document collection. For example, an entity like the Police, might be strongly associated with *Law, Security, Order, and Protection*. However, other entities, like the ANA, ABP, and Security Guards should also be associated with those concepts. That other entities are associated to the same concepts dilutes the salience of the ANP's relationship to the concept. As a result, when one thinks of the police after reading the documents in the collection, no particular aspect of the entity shapes the reader's impression of them, because no aspect is salient.

On the other hand, the Police's association to a concept like *Criminal* may be equally as strong as the other conceptual associations that it has. However, if it far exceeds the association that the ABP and Security Guards, have with *Criminal*, the relationship between the Police and *Criminal* becomes highly salient to the reader. We hypothesize that the increased salience of an association makes it come immediately to mind, and as a result, it becomes the association the reader perceives as the important characteristic of an entity worth the most attention.

In the document collection used here, we examined the extent to which the ANP was associated with five concepts. Of them, TRAINING dominated the associations in the sense that, the ANP had a stronger association to TRAINING than just about any of the entities sampled for our baseline. We suspect that the high degree to which TRAINING has such a salient relationship to the ANP, will bias the reader toward focusing on that aspect of the ANP at the expense of other, equally important, but less salient concepts. In other words, readers may use an *Availability Heuristic* to form their impressions of entities they learn about. That is, a person's notion of what are the important qualities of an entity to focus upon is shaped by the salience of the qualities associated with the entity in the available information. From our reading of the documents discussing training, the dominant sentiment expressed is a positive one in which the reports discuss the work by NATO-led police trainers to create a professional police force. We suspect that, if the reader's impression of the ANP is driven mainly by its association to TRAINING, his/her view of the ANP would follow the positive tone of the reports.

We realize that our discussion about how associative salience drives interpretation is speculative, and requires empirical research. However, the results here have provided a good basis upon which to initiate further work.

How well did GOSSIP do?

GOSSIP's original intended use was as an aid for Intelligence Analysis, in which the connections among entities can provide valuable insights into the structure of social organizations discussed in a document collection. Its use in the Influence Activities/Operations context saw it as a tool for gaining insights into how the impression of entities discussed in a collection of text might be shaped by the way in which the information about entities is presented. The ability to generate profiles of entities was intended to be a secondary capability; however, it became the focus of the study reported here. In what follows, we discuss some of the strengths and weakness of the tool for this purpose.

Strengths. Among its strengths is GOSSIP's ability to extract semantic information about entities in a short time. Within a few seconds, GOSSIP provides the user with a clear notion of which concepts are salient to the entities under examination. This has the potential to save a great amount of time reading documents to find the same information.

Areas for improvement. GOSSIP is still a prototype; hence, we have no doubt that there is room for improvement. However, this study highlighted some basic functionality that could be improved in future versions of the tool.

1. Currently, the documents that connect entities are not sorted according to the relevance they have to a concept. Currently, GOSSIP displays all the documents connecting two individuals regardless of what user-defined concept is of interest to the analyst. So, for

example, if the analyst is interested in the connection between ANP and the entity Kandahar and the concept SAFETY has been activated, it would be useful to see documents connecting the two entities ordered by relevance to the concept. In the study we conducted here, finding examples of documents containing information relevant to concepts required us to search the documents manually. Having GOSSIP sort documents automatically would preclude the need to do so.

2. Currently, the data contained in an entity profile contain only information about how much stronger the association between an entity and concept is relative to a baseline. It may be useful to include the actual cosine between them in a separate table.
3. The tool should include more information about source, and potentially user-defined ratings of reliability that could be applied to various sources.

Overall impression of the tools. GOSSIP has potential to be useful when the analyst needs to gain an understanding of high volumes of information rapidly. For example, when arriving in a theatre of operations, a new analyst must get up to speed on a large amount of information so that he/she can carry on the work of his/her predecessor. GOSSIP can be used effectively as a tool to access the necessary and relevant corporate knowledge required for the analytical task. While in theatre, GOSSIP can be used to track the relationship between entities and concepts of interest over time as a means of measuring campaign effectiveness.

Other uses capitalize on GOSSIP's ability to characterize what are the salient topics in a collection. First, one could use it to help decide the context in which a new story will be presented. For example, if one wished to disseminate a particular message, one might create a message that takes the form of those most popular to the local population. Alternatively, GOSSIP could be used in a similar fashion as a tool for choosing which locally read documents are good candidates for injecting messages to either shape the narratives of the documents or to use the documents as "carriers" for a message.

One point worth mentioning about its use, however, is that information that the tool provides is only as good as the data it ingests. For example, the documents used in this study were selected for a specific audience from open source media that are necessarily edited to lack details that may be important to support accurate analysis. The analytical ability afforded by GOSSIP and other such tools is driven in large part by the text it processes. As such, any understanding that an analyst develops from a document collection will be shaped by the perspective present in the materials. With this in mind, GOSSIP can ingest several document collections, thus allowing the user to compare his/her understanding of a domain from several disparate perspectives. For example, a fuller understanding of the ANP could be examined using several document collections from open source media, local reports, intelligence reports, and situation reports to form an understanding based on several perspective.

Future Directions for GOSSIP as a Tool for Influence Operations

This trial brought to light a number of issues around the use of GOSSIP in theatre. As a device for gaining situational awareness it has its greatest utility as a tool at the start of a rotation for incoming analysts so that they can gain a rapid understanding of the relationships among entities that were learned on the previous rotation. In other words, GOSSIP would be extremely useful in the process of transferring knowledge from one team to the next. During the course of a mission, however, its usefulness depends on the ability to feed the system with data collected from

multiple, often classified, sources from several networks. We did not have this ability during the current trial, and as a result our ability to examine GOSSIP's full capability could not be examined. Based on feedback from CF personnel in theatre, future work for GOSSIP must include:

Making GOSSIP a web-based application which would allow users to gain access to the system from any machine connected to the network using a common web browser, and

Creating a utility that scrapes documents from the network upon which GOSSIP is installed and formatting them properly so that new information can be examined on a daily/weekly basis without having to enter the information manually. Indeed, the requirement for the user to manually enter documents into the system is currently the greatest deterrent for its use.

Conclusions and Recommendations

In this report, we describe the results of a study in which GOSSIP was used as a tool to help uncover the general tone of documents discussing the ANP. The tool provided some useful insights into how certain concepts were associated with the organization, and the subsequent, more qualitative, analysis provided more specific information about how the concepts were being used in the text. In all fairness, however, GOSSIP does not represent the most appropriate tool for all ongoing operations for Influence Activities. Because much of what Influence Activities requires is knowledge about local opinion and perceptions about various organizations and topics, a tool for Opinion Mining and Sentiment Analysis is far more appropriate.

Nevertheless, GOSSIP is a useful tool. A senior analyst in theatre mentioned that GOSSIP would have been useful at the beginning of the operation as a device for uncovering the salience of various concepts discussed in local media. Once the salient concepts are known, articles containing them could be used as the "courier" for the messages that influence operations personnel want to disseminate by adding them to the articles as injects. GOSSIP has some shortcomings, but they can be overcome, and to be fair, the system as it was used in this study is still a prototype. We recommend further development of the tool. As well, we recommend that future opportunities to trial it be exploited to their fullest.

References

Kwantes, P.J. & Terhaar, P. (2010). Entity profiling for intelligence using the Graphical Overview of Social and Semantic Interactions of People (GOSSIP) software tool. *DRDC Technical Report, TR 2010-188*.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

This page intentionally left blank.

List of symbols/abbreviations/acronyms/initialisms

| | |
|--------|--|
| TAV | Technical Assistance Visit |
| IMDB | Internet Movie Database |
| DRDC | Defence Research & Development Canada |
| GOSSIP | Graphical Overview of the Social and Semantic Interactions of People |
| KAF | Kandahar Airfield |
| PSYOPS | Psychological Operations |
| ISAF | International Security Assistance Force |
| ANP | Afghan National Police |
| ANA | Afghan National Army |
| ABP | Afghan Border Police |
| CF | Canadian Forces |
| LSA | Latent Semantic Analysis |
| NATO | North Atlantic Treaty Organization |
| TFK | Task Force Kandahar |
| OMLT | Operational Mentor and Liaison Team |

UNCLASSIFIED

| DOCUMENT CONTROL DATA (Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified) | | |
|---|---|--|
| 1. ORIGINATOR (The name and address of the organization preparing the document, Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's document, or tasking agency, are entered in section 8.) Publishing: DRDC Toronto Performing: DRDC Toronto Monitoring: Contracting: DRDC Toronto | | 2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED |
| 3. TITLE (The complete document title as indicated on the title page. Its classification is indicated by the appropriate abbreviation (S, C, R, or U) in parenthesis at the end of the title) Using Profiles in GOSSIP to Examine Concepts Associated with the Afghan National Police in Open Source Media (U) Utilisation de profils dans le GOSSIP afin d'examiner les concepts reliés à la Police nationale afghane dans les médias : (U) | | |
| 4. AUTHORS (First name, middle initial and last name. If military, show rank, e.g. Maj. John E. Doe.) Peter J Kwantes; Ian Lawless | | |
| 5. DATE OF PUBLICATION (Month and year of publication of document.) August 2011 | 6a NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 14 | 6b. NO. OF REFS (Total cited in document.) 2 |
| 7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of document, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.) Technical Memorandum | | |
| 8. SPONSORING ACTIVITY (The names of the department project office or laboratory sponsoring the research and development – include address.) Sponsoring: DRDC Toronto Tasking: DRDC Toronto | | |
| 9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant under which the document was written. Please specify whether project or grant.) 15ah02 | | 9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.) |
| 10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document) DRDC Toronto TM 2011-085 | | 10b. OTHER DOCUMENT NO(s). (Any other numbers under which may be assigned this document either by the originator or by the sponsor.) |
| 11. DOCUMENT AVAILABILITY (Any limitations on the dissemination of the document, other than those imposed by security classification.) Unlimited distribution | | |
| 12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, when further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.) Unlimited announcement | | |

UNCLASSIFIED

UNCLASSIFIED

DOCUMENT CONTROL DATA

(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

13. **ABSTRACT** (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U) The Graphical Overview of the Social and Semantic Interactions of People (GOSSIP) is a software tool developed by Defence Research and Development Canada – Toronto (DRDC Toronto). The program is designed to help the operator/analyst develop a fast and accurate impression of the relationships among entities (people, places, organizations) discussed in document collections that are too large to read through in a reasonable amount of time. Over the past few years, NATO countries have developed and delivered a training program for Afghan National Police (ANP) members. The ANP are considered by some to be unprofessional, inept, corrupt, and as having little positive effect on the local population's sense of security in their communities. These qualities of the ANP were explored using GOSSIP by examining an open source media collection prepared for analysts and commanders in KAF. We found that when the ANP was discussed in the media updates provided to TFK commanders, it was very often in a positive way. In particular, discussion about NATO's role in professionalizing the ANP dominated articles about the ANP. We propose that the extent to which the ANP is discussed in positively toned articles might lead the target audience for these articles to have an unduly positive impression of the ANP. GOSSIP is a prototype. It needs to be developed further to allow it to be used as a web-based device on a network. Future work should also enhance the tool by providing it with the ability to scrape information from various sources without the user having to load documents manually.

(U) Recherche et développement pour la défense Canada – Toronto (RDDC Toronto) a conçu l'outil logiciel d'aperçu graphique de l'interaction sociale et sémantique entre les personnes (GOSSIP). Ce programme permet à l'opérateur ou l'analyste d'avoir un aperçu rapide et exact des dynamiques interrelationnelles (personnes, places, organismes) présentées dans des documents ne pouvant être lus dans des délais raisonnables en raison de leur volume. Au cours des dernières années, les pays membres de l'Organisation du Traité de l'Atlantique Nord (OTAN) ont mis sur pied un programme d'instruction pour la Police nationale afghane (PNA). Certaines personnes estiment que les membres de la PNA manquent de professionnalisme en plus d'être déplacés et corrompus, et qu'ils apportent un faible sentiment de sécurité à la population locale. Les analystes et commandants de l'aérodrome de Kandahar (KAF) ont examiné, à l'aide du GOSSIP, ces particularités à partir d'une médiathèque ouverte constituée spécialement pour eux. Or, il s'avère que les dernières nouvelles médiatiques sur la PNA qui ont été remises aux commandants de la Force opérationnelle de Kandahar (FO Kandahar) étaient très positives. Les articles traitaient notamment du rôle de l'OTAN en ce qui a trait au professionnalisme de la police afghane. Ainsi, des articles favorables à la PNA pourraient amener le public cible à avoir une meilleure opinion d'elle. Le GOSSIP est un prototype. Il faut le développer davantage afin qu'il devienne un outil Web pouvant être utilisé sur un réseau. Avec le temps, il pourra également regrouper des données de diverses sources sans que l'utilisateur ait à télécharger les documents manuellement.

14. **KEYWORDS, DESCRIPTORS or IDENTIFIERS** (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U)

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale



www.drdc-rddc.gc.ca

